

WHAT IS CLAIMED IS:

- SUB
H
- 5
1. A method for classifying chromatograms, comprising:
receiving a first chromatogram data;
adjusting the first chromatogram data;
reducing the first chromatogram data to a first data set, wherein the first chromatogram data is reduced based upon a consistent positioning across chromatograms;
receiving a second chromatogram data;
adjusting the second chromatogram data;
reducing the second chromatogram data to a second data set, wherein the second chromatogram data is reduced based upon the consistent positioning across chromatograms; and
comparing the first data set and the second data set.
- 10
2. The method of claim 1 in which the acts of adjusting the first and second chromatogram data comprise baseline correction.
- 15
3. The method of claim 1 further comprising:
identifying a first chromatogram region of interest in the first chromatogram data;
identifying a second chromatogram region of interest in the second chromatogram data.
- 20
4. The method of claim 1 in which the act of adjusting data in the first and second regions of interest comprises centering an analysis window around one or more trace features in a given region of interest.
- 25
5. The method of claim 4 in which the act of centering comprises:
determining an average time for the given region of interest; and
centering the analysis window around the average time.
- 30
6. The method of claim 1 further comprising:
filtering the first and second chromatogram data to identify bad data.
7. The method of claim 6 in which the act of filtering is based upon criteria selected a group consisting of: peak height, peak area, peak shape, peak position, peak slope, peak size.

8. The method of claim 1 in which the acts of reducing the first and second chromatogram data to the first and second data sets comprise determining arrays of data set values directly from the first and second chromatogram data.

5 9. The method of claim 1 in which the acts of reducing the first and second chromatogram data to the first and second data sets comprise:

determining an integral of the first and second chromatogram data and plotting against a time axis;

determining a set of time points; and

10 forming arrays of data set values based upon the set of time points and corresponding integral values for the set of time points.

10. The method of claim 1 further comprising:
selecting a reference chromatogram.

11. The method of claim 10 in which the reference chromatogram is selected based upon first selecting a plurality of chromatograms having one or more identified characteristics that most closely match one or more reference characteristics, and identifying a single chromatogram within the plurality of chromatograms to be the reference chromatogram.

12. The method of claim 10 in which other chromatograms are compared against the reference chromatogram.

13. The method of claim 12 further comprising:
25 mapping results of comparing the reference chromatogram against the other chromatograms.

14. The method of claim 13 in which mapping is performed to a two-dimensional cluster map.

30 15. The method of claim 1 in which the act of comparing comprises determining a degree of similarity between the first and second data sets.

16. The method of claim 1 in which the act of comparing comprises determining a degree of dissimilarity between the first and second data sets.

17. The method of claim 1 in which the act of comparing comprises determining distance
5 between vectors associated with the first and second data sets.

18. The method of claim 1 in which the first and second chromatogram data relate to DNA analysis, wherein the reduced chromatogram data excludes a main DNA peak and fully encapsulate a possible sequence variation peak.

10
19. A system for classifying chromatograms, comprising:
a data storage device to store chromatogram data;
a communications interface adaptable to receive chromatogram data from the data
storage device;
15 a data adjustment module to adjust the chromatogram data;
a reduction module to reduce the chromatogram data to a data set that can be compared
against other chromatogram data sets; and
a comparison module to compare the data set against the other chromatogram data sets.

20
20. The system of claim 19 further comprising a bad data filter.

21. The system of claim 20 in which the bad data filter performs filtering based upon criteria selected from the group consisting of: peak height, peak area, peak shape, peak position, peak slope, peak size.

25
22. The system of claim 19 in which the data adjustment module performs baseline correction for the chromatogram data in a region of interest.

23. The system of claim 19 in which the data adjustment module centers the analysis
30 window around one or more trace features in a region of interest.

24. The system of claim 19 in which the reduction module determines an array of data set values directly from the chromatogram data.

25. The system of claim 24 in which the array of data set values are formed by:
selecting a set of time points in the first and second chromatogram data;
determining amplitude values corresponding to the set of time points; and
5 forming the arrays of data set values based upon the set of time points and their
corresponding amplitude values.

26. The system of claim 19 in which the reduction module determines an array of data set
values based upon:
10 determining an integral of the chromatogram data and plotting against a time axis;
determining a set of time points;
forming the arrays of data set values based upon the set of time points and corresponding
integral values for the set of time points.

27. The system of claim 19 implemented using one or more programmable logic devices.

28. The system of claim 19 further comprising a mapping module to map results from the
comparison module.

29. The system of claim 28 further comprising a user interface to display results from the
comparison module.

30. A method for automated classification of chromatograms, comprising:
collecting a first chromatogram data;
25 identifying a first qualitative characteristic for the first chromatogram data;
collecting a second chromatogram data;
identifying a second qualitative characteristic for the second chromatogram data; and
automated comparison of the first qualitative characteristic of the first chromatogram
data to the second qualitative characteristic of the second chromatogram data to classify
30 chromatograms.

31. The method of claim 30 in which the first and second qualitative characteristics of the
first and second chromatogram data comprises peak shape.

32. The method of claim 30 further comprising:
adjusting data from the first chromatogram;
reducing the first chromatogram data to a first data set, wherein the first chromatogram is
reduced based a consistent positioning across chromatograms;
5 adjusting data from the second chromatogram;
reducing the second chromatogram data to a second data set, wherein the second
chromatogram data is reduced based upon the consistent positioning across chromatograms.

33. The method of claim 32 in which the acts of adjusting data from the first and second
10 chromatograms comprise re-centering an analysis window.

34. The method of claim 30 further comprising:
baseline correction of the first and second chromatogram data.

35. The method of claim 30 in which the first and second chromatograms comprise DHPLC
chromatograms.

36. A method for classifying chromatograms, comprising:
receiving a first DHPLC chromatogram;
receiving a second DHPLC chromatogram;
20 automated comparison between the first DHPLC chromatogram and the second DHPLC
chromatogram; and
classifying the first DHPLC chromatograms based upon the results of the automated
comparison.

37. The method of claim 36 wherein the automated comparison is based upon qualitative
analysis.

38. The method of claim 37 in which the qualitative analysis comprises analysis of peak
30 shape.

39. The method of claim 36 in which the first DHPLC chromatogram is classified based
upon likelihood of SNP in DNA corresponding to the first DHPLC chromatogram.

40. The method of claim 36 further comprising:
mapping the first and second chromatogram data.

41. The method of claim 36 further comprising:

- 5 adjusting data from the first DHPLC chromatogram;
reducing the first DHPLC chromatogram data to a first data set, wherein the first DHPLC
chromatogram is reduced based a consistent positioning across chromatograms;
adjusting data from the second DHPLC chromatogram;
reducing the second DHPLC chromatogram data to a second data set, wherein the second
10 chromatogram DHPLC data is reduced based upon the consistent positioning across
chromatograms.

42. A user interface for displaying results of chromatogram analysis, comprising:
a cluster map display portion;
a raw chromatogram display portion; and
a normalized chromatogram display portion.

43. A computer usable medium having stored thereon a sequence of instructions which,
when executed by a processor, causes the processor to execute a process for classifying
chromatograms, said process comprising:

- 20 receiving a first chromatogram data;
adjusting the first chromatogram data;
reducing the first chromatogram data to a first data set, wherein the first chromatogram
data is reduced based a consistent positioning across chromatograms;
25 receiving a second chromatogram data;
adjusting the second chromatogram data;
reducing the second chromatogram data to a second data set, wherein the second
chromatogram data is reduced based upon the consistent positioning across chromatograms; and
comparing the first data set and the second data set.